# Time Majority Voting, a PC-based EEG Classifier for Non-expert Users

Guangyao Dou, Zheng Zhou, Xiaodong Qu

Brandeis University, Swarthmore College

Human-Computer Interaction International 2022

# Introduction

Electroencephalogram(EEG) - A recording of electrical activity in the brain from the scalp surface areas.

Current Challenges:
- Small Data size
- Low signal-to-noise ratio
- Outlier detection

# Outline

- Motivation
- Methods
- Dataset
- Experiments
- Results
- Improvements
- Conclusion

# Motivation

Current machine learnings are more suitable for EEG data collected in clinical settings.

We want to make it easier for users to analyze EEG data when subjects collect their EEG data in non-clinical settings, such as at home.

# Methods

State-of-the-art Machine Learning and Deep Learning algorithms:

- Support Vector Machine with RBF/linear kernel function (RBF/linear SVM)
- Linear Discriminant Analysis
- K-Nearest Neighbors
- Decision Trees
- Ensemble methods: Random Forest, AdaBoost, GradientBoost, etc
- Deep Learning: CNN, RNN, especially LSTM
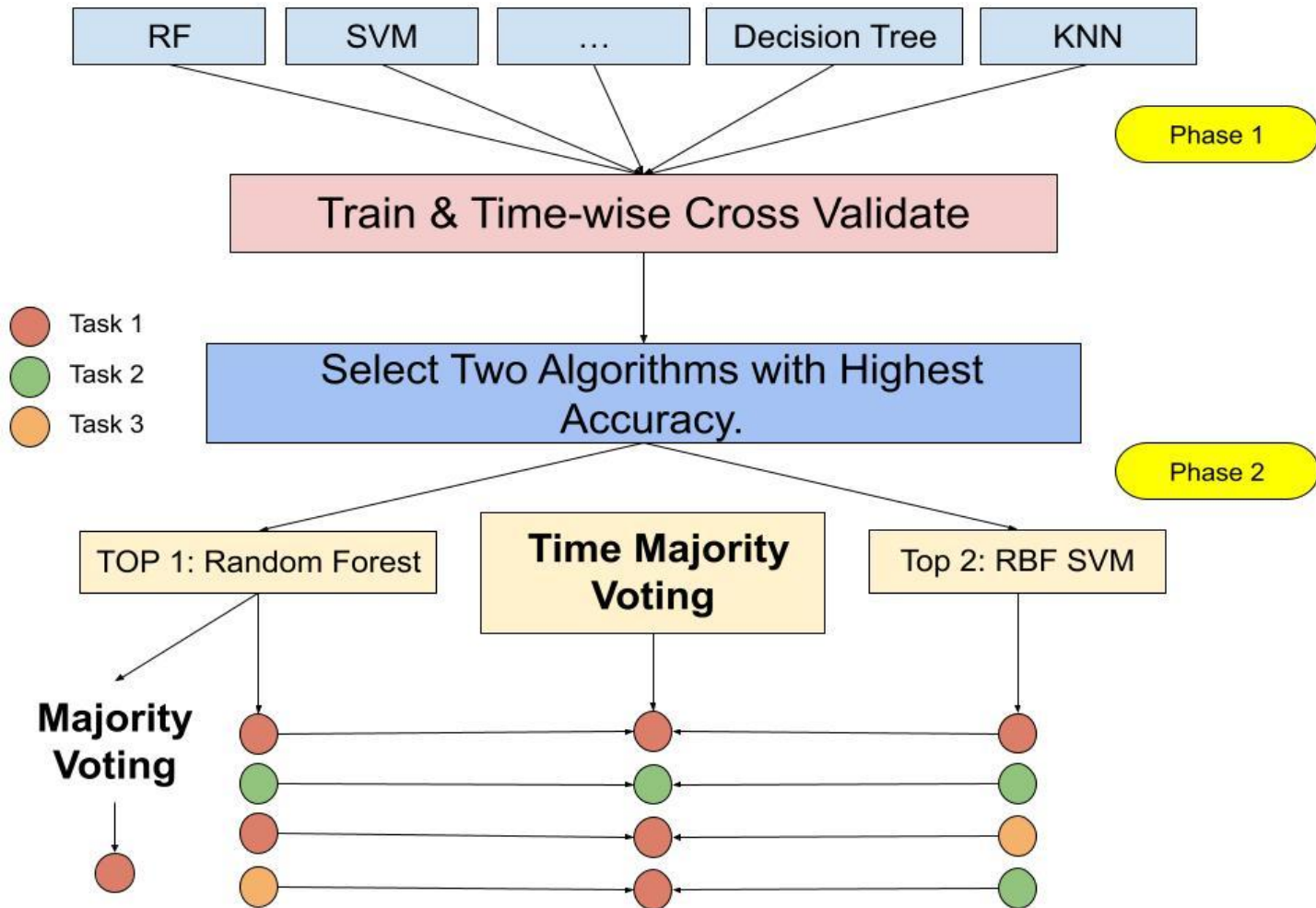
# Data

-   A dataset with 17 subjects in total
-   There were six sessions for each subject
-   There were five tasks (Think, Count, Recall, Draw, Breathe) performed in each session
-   Each task lasted for one minute.
-   More details: Qu, X., Liu, P., Li, Z., & Hickey, T. (2020, October). Multi-class time continuity voting for EEG classification. In *International Conference on Brain Function Assessment in Learning* (pp. 24-33). Springer, Cham.
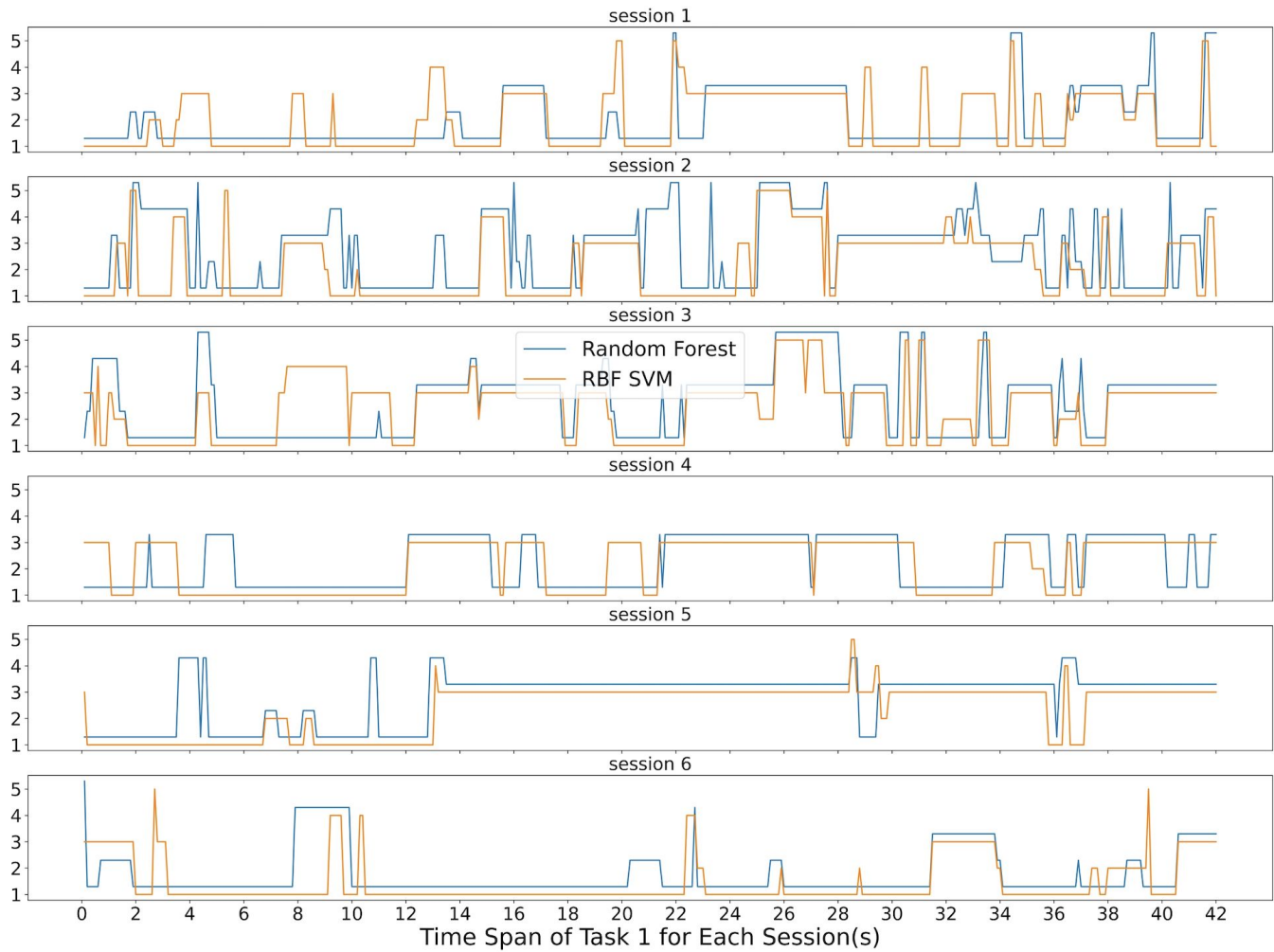
# Experiments

- Data cleaning
  - The first 18 seconds will be removed for each task in each session
  - Anomaly detected for consecutive 1.4 seconds
  - Exclude subjects that lost more than 65% of data
  - Exclude sessions that lost more than 65% of data for each subject
- Time-wise cross-validation
  - Split data evenly and continuously into seven folds
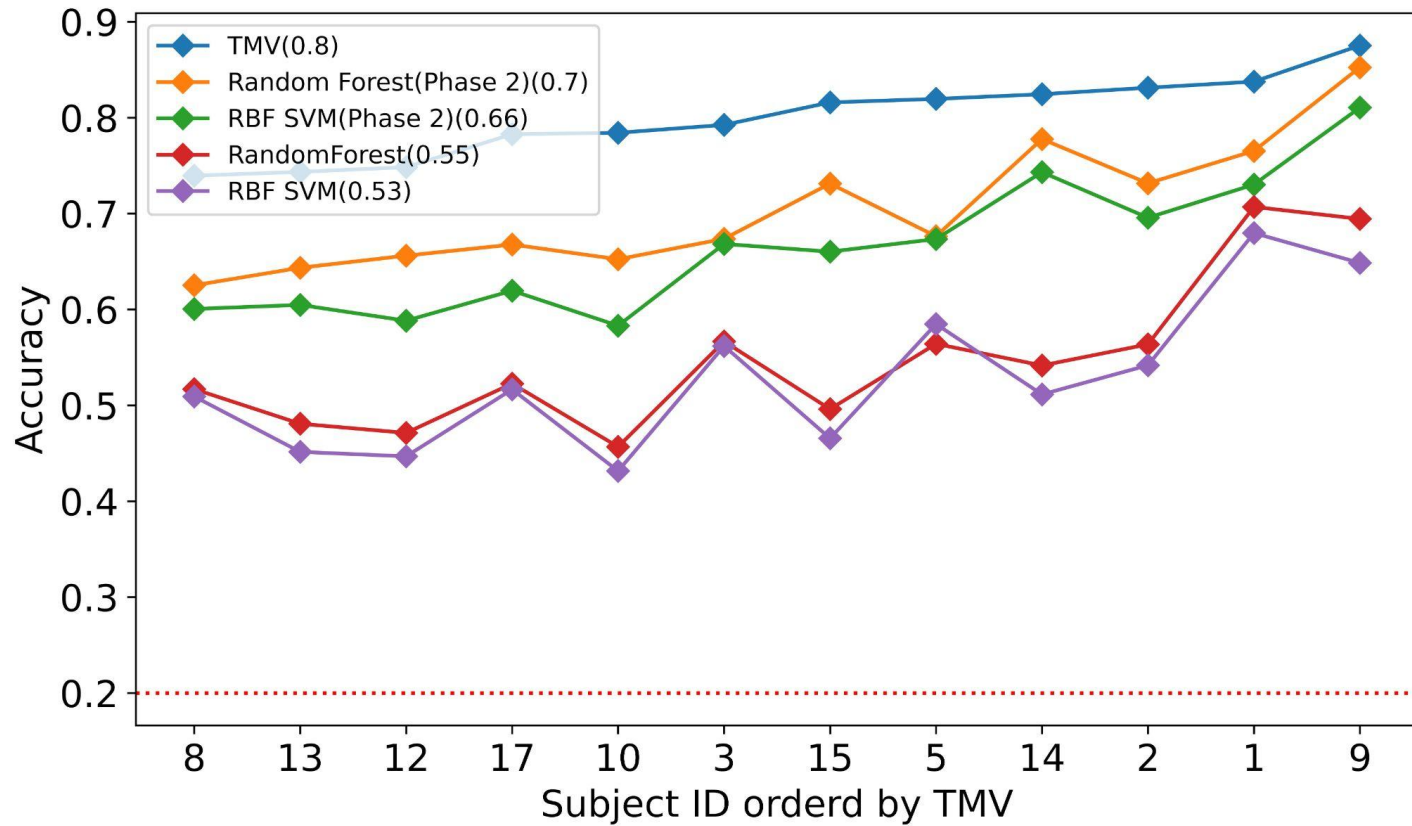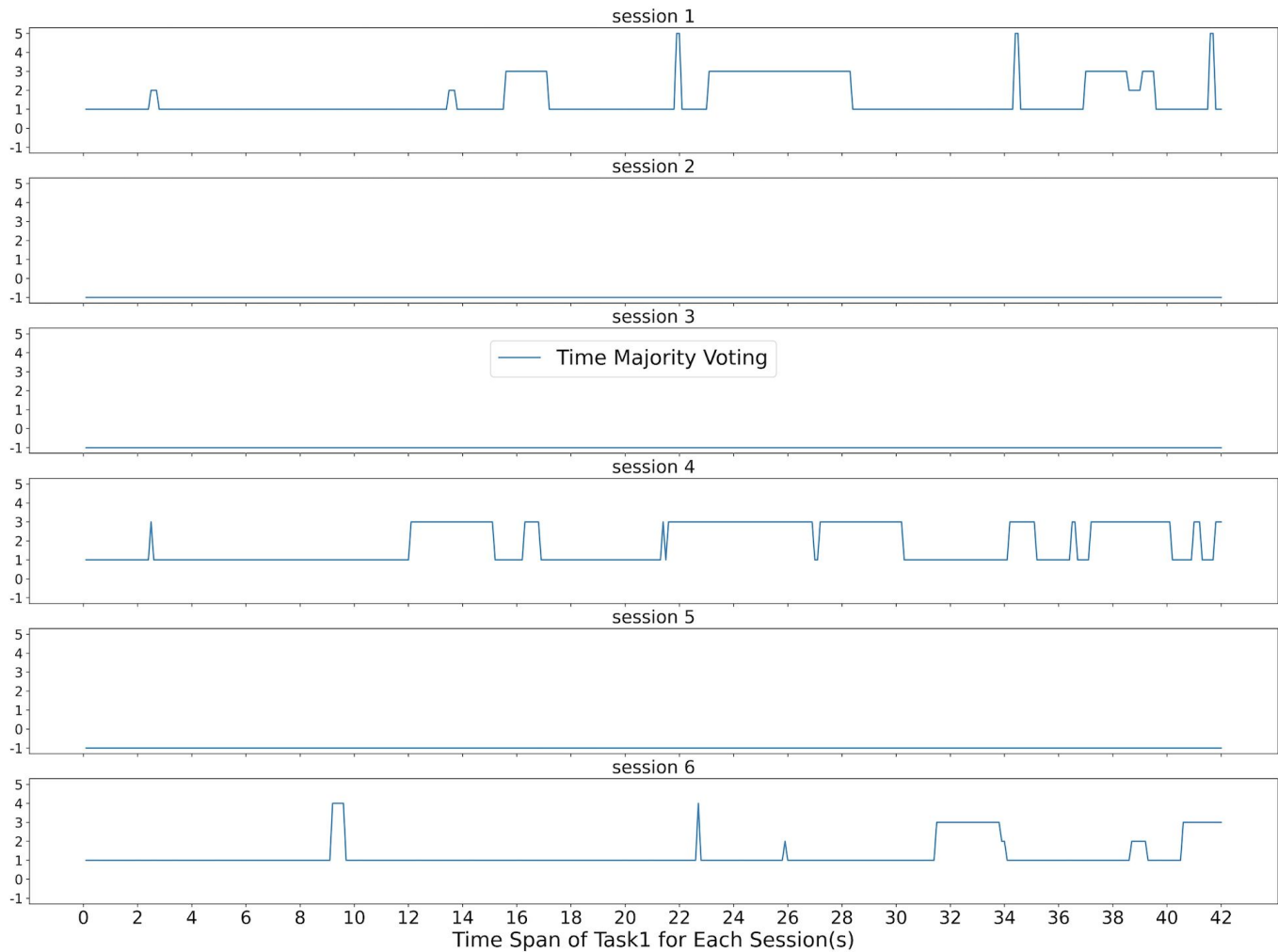  - Exclude folds that lost more than 65% of data

# Results
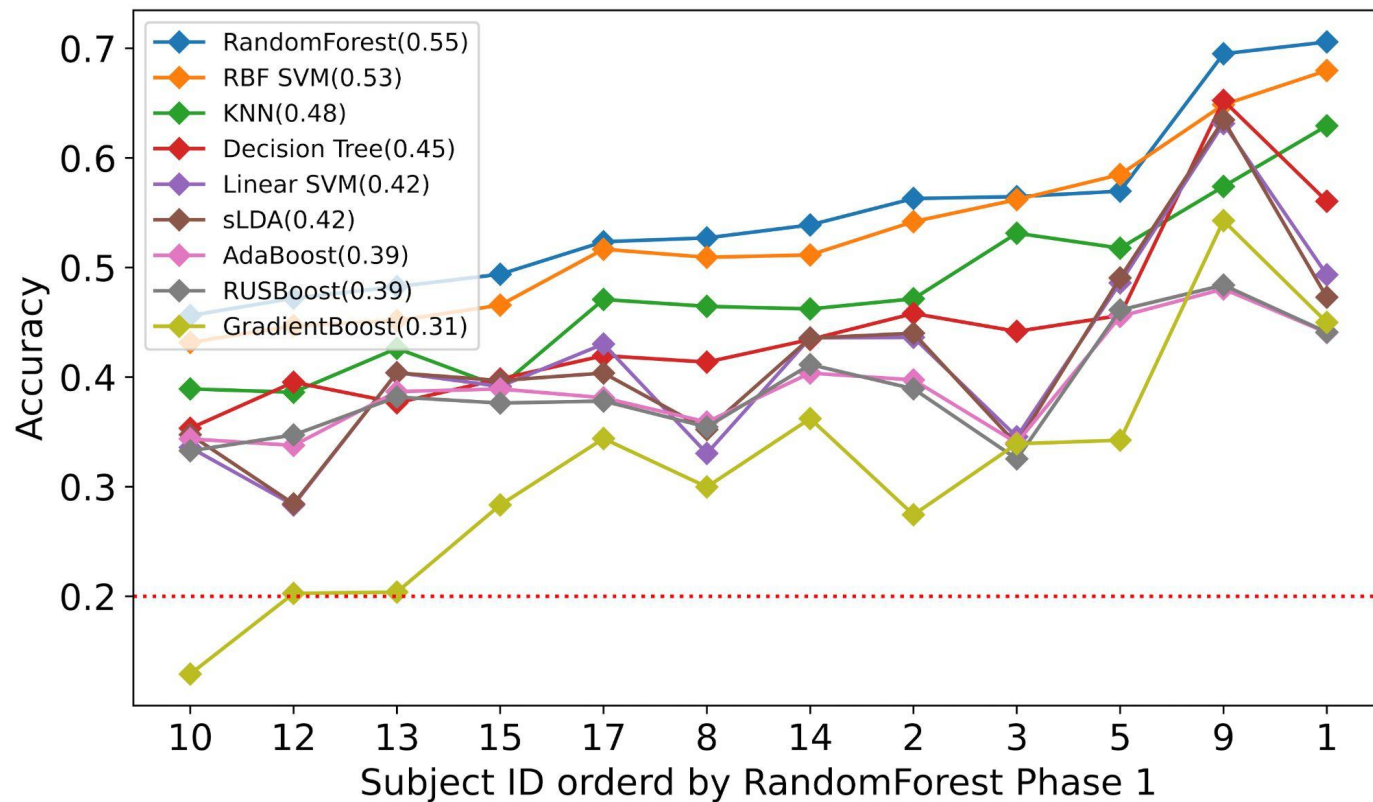
| Algorithms | Average Accuracy | Average code run-time (s) |
|---|---|---|
| Random Forest Phase 1 | **0.55** | 42.0 |
| RBF SVM | **0.53** | 30.5 |
| Nearest Neighbors | 0.48 | **1.9** |
| Decision Tree | 0.44 | 0.9 |
| Linear SVM | 0.42 | 23.2 |
| Shrinkage LDA | 0.42 | 0.2 |
| Adaboost Classifier | 0.39 | 47.8 |
| RUSBoost | 0.39 | 28.2 |
| GradientBoost | 0.31 | 24.0 |

**Table 2.** State-of-the-art Algorithms with Accuracy and Run-time

session 1

session 2

session 3

session 4

session 5

session 6

Time Span of Task 1 for Each Session(s)

Random Forest
RBF SVM

session 1

session 2

session 3

Time Majority Voting

session 4

session 5

session 6

Time Span of Task1 for Each Session(s)
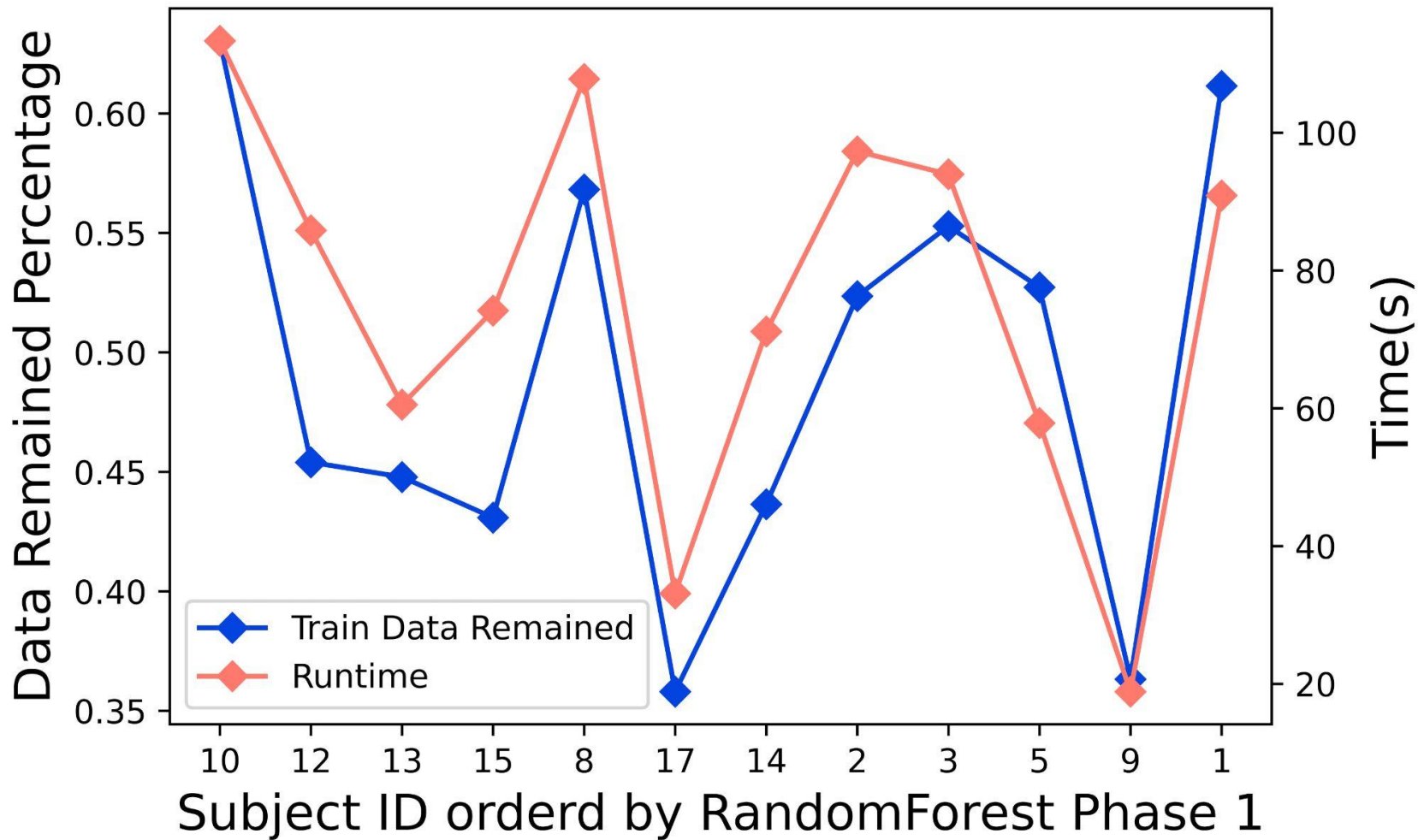
# Future Direction

- Perform voting with more classifiers
- Remove outliers and noises
- Test on the larger dataset

# Conclusion

- The results demonstrated that TMV outperformed other state-of-the-art existing classifiers.
- The interpretability of TMV can contribute to a better understanding of the machine learning analysis and an improved design for future experiments.

Thank you so much!

Contact us: guangyaodou@brandeis.edu

Questions?